Monte Carlo Artificial Intelligence: Bayesian Networks

Why This Matters

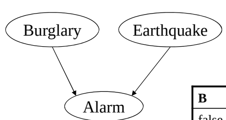
- Bayesian networks have been the most important contribution to the field of AI in the last 10 years
- Provide a way to represent knowledge in an uncertain domain and a way to reason about this knowledge
- Many applications: medicine, factories, help desks, spam filtering, etc.

A Bayesian Network

B	P(B)
false	0.999
true	0.001

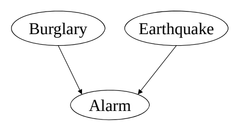
E	P(E)
false	0.998
true	0.002

A Bayesian network is made up of two parts:
1. A directed acyclic graph
2. A set of parameters



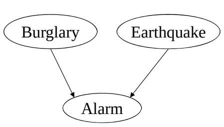
B	E	A	P(A|B,E)
false	false	false	0.999
false	false	true	0.001
false	true	false	0.71
false	true	true	0.29
true	false	false	0.06
true	false	true	0.94
true	true	false	0.05
true	true	true	0.95

A Directed Acyclic Graph



1. A directed acyclic graph:
- The nodes are random variables (which can be discrete or continuous)
- Arrows connect pairs of nodes (X is a parent of Y if there is an arrow from node X to node Y).

A Directed Acyclic Graph

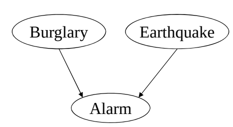


- Intuitively, an arrow from node X to node Y means X has a direct influence on Y (we can say X has a casual effect on Y)
- Easy for a domain expert to determine these relationships
- The absence/presence of arrows will be made more precise later on

A Set of Parameters

B	P(B)
false	0.999
true	0.001

E	P(E)
false	0.998
true	0.002



B	E	A	P(A|B,E)
false	false	false	0.999
false	false	true	0.001
false	true	false	0.71
false	true	true	0.29
true	false	false	0.06
true	false	true	0.94
true	true	false	0.05
true	true	true	0.95

Each node X_i has a conditional probability distribution $P(X_i \mid Parents(X_i))$ that quantifies the effect of the parents on the node

The parameters are the probabilities in these conditional probability distributions

Because we have discrete random variables, we have conditional probability tables (CPTs)

A Set of Parameters

Conditional Probability Distribution for Alarm

B	E	A	P(A|B,E)
false	false	false	0.999
false	false	true	0.001
false	true	false	0.71
false	true	true	0.29
true	false	false	0.06
true	false	true	0.94
true	true	false	0.05
true	true	true	0.95

Stores the probability distribution for Alarm given the values of Burglary and Earthquake

For a given combination of values of the parents (B and E in this example), the entries for P(A=true|B,E) and P(A=false|B,E) must add up to 1 eg. P(A=true|B=false,E=false) + P(A=false|B=false,E=false)=1

If you have a Boolean variable with k Boolean parents, how big is the conditional probability table?

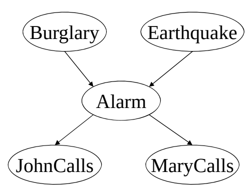
How many entries are independently specifiable?

A Representation of the Full Joint Distribution

- We will use the following notation:
  - *parents*(X_i) for the values of the parents of X_i
- From the Bayes net, we can calculate:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

Example



P(*JohnCalls*, *MaryCalls*, *Alarm*, *Burglary*, *Earthquake*)

= ***P***(*JohnCalls* | *Alarm*) ***P***(*MaryCalls* | *Alarm*) ***P***(*Alarm* | *Burglary*, *Earthquake*) ***P***(*Burglary*) ***P***(*Earthquake*)

Conditional Independence

We can look at the actual graph structure and determine conditional independence relationships eg.

1. A node (*X*) is conditionally independent of its non-descendants (Z_{1j}, Z_{nj}), given its parents (U_1, U_m).